IT & C

ISSN 2821 - 8469, ISSN – L 2821 - 8469, Volumul 2, Numărul 3, Septembrie 2023

Provocări în inteligența artificială

Nicolae Sfetcu

Sfetcu, Nicolae (2023), Provocări în inteligența artificială, *IT & C*, **2**:3, 3-10, DOI: 10.58679/IT24537, https://www.internetmobile.ro/provocari-in-inteligenta-artificiala/

Publicat online: 05.08.2023

Provocări în inteligența artificială

Nicolae Sfetcu[1]
nicolae@sfetcu.com

Challenges in artificial intelligence

Abstract

Artificial intelligence is a transformative field that has captured the attention of scientists, engineers, businesses and governments around the world. As we move further into the 21st century, some prominent trends in AI have emerged. Artificial intelligence and machine learning technology are used in most of the essential applications of the 2020s.

AI capability control proposals, also more restrictively referred to as AI confinement, aim to increase the ability to monitor and control the behavior of AI systems, including artificial general intelligence, proposed to reduce the danger they might pose if misaligned.

The researches and development aimed at validating answers to all these questions must include philosophy for its robust formalisms.

Keywords: artificial intelligence, challenges, risks, trends, deep learning, neural networks, ethical, superintelligence, transhumanism, artificial general intelligence

Rezumat

Inteligența artificială este un domeniu transformator care a captat atenția oamenilor de știință, inginerilor, întreprinderilor și guvernelor din întreaga lume. Pe măsură ce avansăm mai departe în secolul 21, au apărut câteva tendințe proeminente în domeniul IA. Inteligența artificială și tehnologia de învățare automată sunt utilizate în majoritatea aplicațiilor esențiale ale anilor 2020.

Propunerile de control al capabilității inteligenței artificiale, denumite și în mod mai restrictiv confinarea IA, urmăresc să sporească posibilitatea de a monitoriza și controla comportamentul sistemelor IA, inclusiv inteligența generală artificială, propusă pentru a reduce pericolul pe care l-ar putea prezenta dacă sunt nealiniate.

[1] Cercetător - Academia Română - Comitetul Român de Istoria și Filosofia Științei și Tehnicii (CRIFST), Divizia de Istoria Științei (DIS)

Cercetarea și dezvoltarea menite să valideze răspunsuri la toate aceste întrebări trebuie să includă filozofia pentru formalismele sale robuste.

Cuvinte cheie: inteligența artificială, provocări, riscuri, tendințe, învățarea profundă, rețele neuronale, etic, superinteligența, transumanism, inteligența generală artificială

IT & C, Volumul 2, Numărul 3, Septembrie 2023, pp. 3-10
ISSN 2821 - 8469, ISSN – L 2821 – 8469, DOI: 10.58679/IT24537
URL: https://www.internetmobile.ro/provocari-in-inteligenta-artificiala/

Introducere

Inteligența artificială (IA) este un domeniu transformator care a captat atenția oamenilor de știință, inginerilor, întreprinderilor și guvernelor din întreaga lume. Creșterea sa rapidă a stârnit o imensă curiozitate și entuziasm, precum și îngrijorări cu privire la impactul său potențial asupra societății. Pe măsură ce avansăm mai departe în secolul 21, au apărut câteva tendințe proeminente în domeniul IA, modelând modul în care interacționăm cu tehnologia, sporindu-ne capacitățile și redefinind diverse industrii.

Inteligența artificială și tehnologia de învățare automată sunt utilizate în majoritatea aplicațiilor esențiale ale anilor 2020, inclusiv: motoarele de căutare (cum ar fi Căutarea Google), direcționarea reclamelor online, sisteme de recomandare (oferite de Netflix, YouTube sau Amazon), generarea traficului pe internet publicitate direcționată (AdSense, Facebook), asistenți virtuali (cum ar fi Siri sau Alexa), vehicule autonome (inclusiv drone, ADAS și mașini cu conducere autonomă), traducere automată a limbii (Microsoft Translator, Google Translate), recunoașterea facială (Apple's Face ID sau Microsoft's DeepFace) și etichetarea imaginilor (utilizată de Facebook, iPhoto de la Apple și TikTok).

La începutul anilor 2020, inteligența artificială generativă a câștigat o proeminență pe scară largă, precum ChatGPT.

Tendințe în inteligența artificială

Învățarea profundă și rețele neuronale

Învățarea profundă, un subset al învățării automate, a devenit o piatră de temelie a progresului IA. Activată de rețelele neuronale, permite algoritmilor să învețe modele complexe din cantități mari de date, realizând sarcini care anterior erau considerate dincolo de sfera mașinilor. Rețelele neuronale convoluționale (CNN) și rețelele neuronale recurente (RNN) au revoluționat viziunea computerizată, procesarea limbajului natural și recunoașterea vorbirii. Rafinarea continuă a acestor arhitecturi a condus la descoperiri în domenii precum sinteza imaginilor, traducerea în timp real și diagnosticarea medicală.

Viziunea computerizată (CV) este ramura IA care se ocupă cu procesarea și înțelegerea informațiilor vizuale, cum ar fi imagini și videoclipuri. CV permite aplicații precum recunoașterea feței, detectarea obiectelor, segmentarea scenei, analiza imaginilor medicale, mașinile cu conducere autonomă și multe altele. CV a avansat odată cu utilizarea rețelelor neuronale convoluționale (CNN), care pot învăța să extragă caracteristici și modele din datele vizuale.

IA în procesarea limbajului natural (NLP)

Procesarea limbajului natural (NLP) este ramura IA care se ocupă cu analiza și generarea limbajului natural, cum ar fi vorbirea și textul. NLP permite aplicații precum chatbot, asistenți vocali, traducere automată, analiză a sentimentelor, rezumat text și multe altele.

Capacitatea IA de a înțelege și genera limbajul uman a înregistrat progrese remarcabile. Aplicațiile procesării limbajului natural (NLP), cum ar fi chatbot, asistenți vocali, traducere automată, analiză a sentimentelor, rezumat text, analiza sentimentelor și rezumarea textului și multe altele, au făcut interacțiunile cu computerele mai naturale și mai eficiente. NLP s-a îmbunătățit odată cu utilizarea modelelor de învățare profundă, cum ar fi transformatoarele, care pot capta informațiile semantice și sintactice ale limbajului natural. Modelele de limbaj pre-antrenate, cum ar fi ChatGPT al OpenAI, au sporit semnificativ performanța diferitelor sarcini NLP. Pe măsură ce modelele lingvistice continuă să evolueze, acestea sunt integrate într-o gamă diversă de produse și servicii, transformând asistența pentru clienți, crearea de conținut și regăsirea informațiilor.

Învățarea prin consolidare și sisteme autonome

Învățarea prin consolidare (RL) este ramura IA care se ocupă cu învățarea din încercare și eroare, bazată pe recompense și penalități, apărând ca o tehnică puternică de instruire a agenților IA pentru a lua decizii în medii dinamice. Sistemele autonome, alimentate de învățare prin consolidare, au câștigat importanță în sectoare precum robotica, mașinile cu conducere autonomă și automatizarea industrială. RL s-a dezvoltat cu ajutorul rețelelor neuronale profunde (DNN), care pot învăța politici și strategii complexe din date cu dimensiuni mari. Aceste tehnologii au potențialul de a spori siguranța, de a crește eficiența și de a revoluționa industriile de transport și logistică.

Rețele adverse generative (GAN) sunt un tip de rețea neuronală care poate genera date realiste și noi, cum ar fi imagini, videoclipuri, text și audio. Aceste rețele constau din două rețele concurente: un generator care încearcă să creeze date false și un discriminator care încearcă să facă distincția între datele reale și cele false. Pot fi utilizate pentru aplicații precum sinteza imaginilor, transferul stilului, super-rezoluția, creșterea datelor și multe altele.

Etica și explicabilitatea IA

Implementarea din ce în ce mai mare a IA a stârnit îngrijorări cu privire la implicațiile etice și problema „cutiei negre”. Pe măsură ce sistemele IA devin mai complexe, înțelegerea proceselor lor de luare a deciziilor devine o provocare, ceea ce duce la potențiale părtiniri și consecințe nedorite. Cercetătorii și factorii de decizie politică lucrează pentru a aborda aceste probleme prin promovarea transparenței, a răspunderii și a dezvoltării de modele IA explicabile. Cadrele etice de inteligență artificială urmăresc să se asigure că tehnologiile de inteligență artificială sunt dezvoltate și utilizate în mod responsabil, respectând valorile și drepturile omului.

Edge AI și învățarea federată

Edge AI implică utilizarea edge computing în inteligența artificială, prin rularea algoritmilor IA pe dispozitive locale, mai degrabă decât să se bazeze doar pe infrastructura bazată pe cloud. Această tendință permite procesarea datelor în timp real, o latență redusă și confidențialitate îmbunătățită, făcând-o ideală pentru aplicații precum dispozitive IoT, asistență medicală și vehicule autonome. Învățarea federată, un subset al Edge AI, permite mai multor

dispozitive să antreneze în colaborare un model IA partajat fără a partaja date brute, păstrând confidențialitatea utilizatorilor, beneficiind în același timp de cunoștințele colective.

IA în asistența medicală

IA a făcut progrese semnificative în sectorul asistenței medicale, asistând profesioniștii din domeniul medical în diagnosticare, planificare a tratamentului, descoperirea medicamentelor și monitorizarea pacienților. Modelele de învățare automată analizează imagini medicale, prezic progresia bolii și identifică modele în datele pacienților pentru a oferi soluții personalizate de asistență medicală. Dispozitivele portabile alimentate cu inteligență artificială și dispozitivele de monitorizare de la distanță au permis urmărirea continuă a sănătății, dând indivizii putere să preia controlul asupra bunăstării lor.

Potențiale riscuri ale inteligenței artificiale

Propunerile de control al capabilității IA, denumite și în mod mai restrictiv confinarea IA, urmăresc să sporească posibilitatea de a monitoriza și controla comportamentul sistemelor IA, inclusiv inteligența generală artificială (AGI) propusă pentru a reduce pericolul pe care l-ar putea prezenta dacă sunt nealiniate (duc la consecințe neintenționate de proiectant). Cu toate acestea, controlul capacității devine mai puțin eficient pe măsură ce agenții devin mai inteligenți și capacitatea lor de a exploata defectele sistemelor de control uman crește, ceea ce poate duce la un risc existențial în cazul AGI. Prin urmare, filozoful de la Oxford Nick Bostrom și alții recomandă metodele de control al capacității doar ca supliment la metodele de aliniere.

Superinteligența și singularitatea

O superinteligență este un agent ipotetic care ar poseda o inteligență care o depășește cu mult pe cea a minții umane cele mai strălucitoare și mai talentate. Dacă cercetarea în inteligența generală artificială ar produce un software suficient de inteligent, acesta ar putea fi capabil să se reprogrameze și să se îmbunătățească. Software-ul îmbunătățit s-ar putea îmbunătăți și mai bine, ducând la ceea ce I. J. Good a numit o „explozie de informații”, iar Vernor Vinge a numit o „singularitate Cu toate acestea, majoritatea tehnologiilor (cum ar fi dele din transport) nu se îmbunătățesc exponențial la nesfârșit, ci mai degrabă urmează o curbă în S, încetinind atunci când ating limitele fizice ale ceea ce poate face tehnologia.

Bostrom a pictat recent o imagine extrem de întunecată a unui posibil viitor. El subliniază că „prima superinteligență” ar putea avea capacitatea de a modela viitorul vieții originare de Pământ, ar putea avea cu ușurință obiective finale non-antropomorfe și ar avea probabil motive instrumentale pentru a urmări achiziția de resurse nelimitată.

Există două cauze independente din punct de vedere logic, dar care se întăresc reciproc, ale îmbunătățirii inteligenței: creșteri ale vitezei de calcul și îmbunătățiri ale algoritmilor utilizați. Prima este prezisă de Legea lui Moore și de îmbunătățirile prognozate în hardware, și este comparativ similară cu progresele tehnologice anterioare. Dar există unii cercetători IA care cred că software-ul este mai important decât hardware-ul. Tehnologi și academicieni proeminenți contestă plauzibilitatea unei singularități tehnologice. Robin Hanson și-a exprimat scepticismul față de creșterea inteligenței umane, scriind că, odată ce „fructul de jos” al metodelor ușoare de creștere a inteligenței umane va fi epuizat, îmbunătățirile ulterioare vor deveni din ce în ce mai dificile.

Riscul existențial

S-a susținut că inteligența artificială va deveni atât de puternică încât umanitatea poate pierde ireversibil controlul asupra acesteia. Acest lucru ar putea, așa cum spune fizicianul Stephen Hawking, „ănsemna sfârșitul rasei umane”. Potrivit filozofului Nick Bostrom, pentru aproape orice obiective pe care le poate avea o IA suficient de inteligentă, este stimulată instrumental să se protejeze de închidere și să dobândească mai multe resurse, ca pași intermediari pentru a atinge mai bine aceste obiective. Sentința sau emoțiile nu sunt necesare pentru ca o IA avansată să fie periculoasă. Pentru a fi în siguranță pentru umanitate, o superinteligență ar trebui să fie aliniată cu adevărat cu morala și valorile umanității, astfel încât să fie „în mod fundamental de partea noastră”. Politologul Charles T. Rubin a susținut că „orice bunăvoință suficient de avansată poate fi imposibil de distins de reavoință” și a avertizat că nu ar trebui să fim încrezători că mașinile inteligente ne vor trata implicit în mod favorabil.

Dacă o mașină superinteligentă dominantă ar concluziona că supraviețuirea umană este un risc inutil sau o risipă de resurse, rezultatul ar fi dispariția umană. Acest lucru s-ar putea întâmpla dacă o mașină, programată fără respect pentru valorile umane, dobândește în mod neașteptat superinteligență prin auto-îmbunătățire recursivă sau reușește să scape din reținerea sa într-un scenariu AI Box.

Dar chiar dacă nu permitem mașinilor să ia decizii, controlul acestor mașini este probabil să fie deținut de o mică elită care va considera restul umanității ca fiind inutilă – deoarece mașinile pot face orice va fi nevoie.

Opiniile dintre experți și din interiorul industriei sunt mixte, cu fracțiuni considerabile atât preocupate, cât și nepreocupate de riscul din eventuala IA superinteligentă. Personalități precum Stephen Hawking, Bill Gates, Elon Musk și-au exprimat îngrijorarea cu privire la riscul existențial din IA. În 2023, pionierii inteligenței artificiale, inclusiv Geoffrey Hinton, Yoshua Bengio, Demis Hassabis și Sam Altman, au emis declarația comună că „atenuarea riscului de dispariție din cauza inteligenței artificiale ar trebui să fie o prioritate globală alături de alte riscuri la scară societală, cum ar fi pandemiile și războiul nuclear"; alții, precum Yann LeCun, consideră că acest lucru este nefondat. Mark Zuckerberg a spus că IA va „debloca o cantitate imensă de lucruri pozitive", inclusiv vindecarea bolilor și îmbunătățirea siguranței mașinilor care se conduc singure. Unii experți au susținut că riscurile sunt prea îndepărtate în viitor pentru a justifica cercetări sau că oamenii vor fi valoroși din perspectiva unei mașini superinteligente. Rodney Brooks, în special, a spus în 2014 că IA „răuvoitoare" este încă la secole distanță.

Transumanismul

Gânditorii transhumaniști studiază potențialele beneficii și pericolele tehnologiilor emergente care ar putea depăși limitările fundamentale ale umanității, precum și etica utilizării unor astfel de tehnologii. Unii transhumaniști consideră că ființele umane ar putea în cele din urmă să se transforme în ființe cu abilități atât de mult extinse din condiția actuală, încât să merite eticheta de ființe postumane.

Designerul de roboți Hans Moravec, ciberneticianul Kevin Warwick și inventatorul Ray Kurzweil au prezis că oamenii și mașinile se vor îmbina în viitor în cyborgi care sunt mai capabili și mai puternici decât oricare dintre ele. Această idee, numită transumanism, își are rădăcinile în Aldous Huxley și Robert Ettinger. Edward Fredkin susține că „inteligenta artificială este următoarea etapă a evoluției", idee propusă pentru prima dată de „Darwin printre mașini" a lui Samuel Butler încă din 1863 și extinsă de George Dyson în cartea sa cu același nume în 1998. Kurzweil susține că în viitor „Nu va exista nicio distincție, post-singularitate, între om și mașină sau între realitatea fizică și cea virtuală". Andy Clark prevede că oamenii vor deveni treptat, cel

puțin într-o măsură apreciabilă, cyborgi, prin amabilitatea membrelor artificiale și a organelor de simț și a implanturilor.

Concluzie

Există câteva lucruri pe care le putem spune în siguranță despre mâine. Cu siguranță, știm acum că IA va reuși să producă ființe artificiale. De fapt, multe locuri de muncă făcute în prezent de oameni vor fi cu siguranță făcute de ființe artificiale programate corespunzător. Daimler rulează deja reclame în care promovează capacitatea mașinilor lor de a conduce „autonom”, permițând ocupanților umani ai acestor vehicule să ignore drumul și să citească. Alte exemple ar include: curățătoriile, poşta, funcționarii de birou, cercetașii militari, chirurgii și piloții. O altă predicție înrudită este că IA va putea juca rolul unei proteze cognitive pentru oameni. Viziunea protezei vede IA ca un „mare egalizator” care ar duce la mai puțină stratificare în societate, probabil similar cu modul în care sistemul numeric hindu-arab a pus aritmetica la dispoziția maselor și cu modul în care presa Guttenberg a contribuit la alfabetizarea care a devenit mai universală.

Cercetarea și dezvoltarea menite să valideze răspunsuri la toate aceste întrebări trebuie să includă filozofia (domeniul la care se apelează pentru formalisme robuste care să modeleze atitudinile propoziționale umane în termenii mașinii).

Inteligența artificială continuă să modeleze lumea noastră în moduri profunde, conducând la inovații în diverse domenii. Tendințele în IA demonstrează potențialul IA de a crea progrese remarcabile, făcându-ne viața mai eficientă, productivă și confortabilă. Cu toate acestea, este esențial să se abordeze provocările etice și potențialele părtiniri asociate cu IA, asigurând dezvoltarea și implementarea responsabilă a acesteia. Pe măsură ce IA continuă să evolueze, este crucial pentru factorii de decizie, cercetători și societate în ansamblu să colaboreze și să navigheze în mod responsabil în viitorul acestei tehnologii transformatoare. Cu abordarea corectă, IA are potențialul de a revoluționa industriile, de a împuternici indivizii și de a depăși granițele progresului uman.